Betting markets for English Premier League results and scorelines: evaluating a forecasting model

Article

Accepted Version

Reade, J., Singleton, C. and Vaughan Williams, L. (2020) Betting markets for English Premier League results and scorelines: evaluating a forecasting model. Economic Issues, 25 (1). pp. 87-106. ISSN 1363-7029 Available at https://centaur.reading.ac.uk/89738/

Betting markets for English Premier League results and scorelines: evaluating a simple forecasting model

J. James Reade
University of Reading

Carl Singleton
University of Reading

Leighton Vaughan Williams*
Nottingham Trent University

March 2020
Link to latest version

Abstract

Using betting odds from two recent seasons of English Premier League football matches, we evaluate probability and point forecasts generated from a standard statistical model of goal scoring. The bookmaker odds show significant evidence of the favourite-longshot bias for exact scorelines, which is not generally present for match results. We find evidence that the scoreline probability forecasts from the model are better than what the odds of bookmakers imply, based on forecast encompassing regressions. However, when we apply a simple betting strategy using point forecasts from the model, there are no substantial or consistent financial returns to be made over the two seasons. In other words, there is no evidence from this particular statistical model that the result, scoreline, margin of victory or total goals betting markets are on average inefficient.

Keywords: Forecasting; Statistical modelling; Regression models; Prediction markets
JEL Codes: C53; G14; G17; L83

*Corresponding author: j.j.reade@reading.ac.uk, Department of Economics, University of Reading, Whiteknights Campus, RG6 6UA, UK.

c.a.singleton@reading.ac.uk, Department of Economics, University of Reading, UK. leighton.vaughan-williams@ntu.ac.uk, Nottingham Business School, Nottingham Trent University, UK.

1 Introduction

In this study, we evaluate two sources of association football (soccer) match forecasts: betting markets and a standard statistical model. Ultimately, the two most important aspects to the outcome of a football match are the *result* and the *scoreline*. The result is a win for either team, or a draw (tie). The scoreline gives the number of goals scored by each team. A football scoreline is a pair of non-negative integers, which may be correlated due to the common conditions both teams face within a match, or because we expect that teams and their tactics will respond within matches to the goals scored (or not) by their opponents (e.g., Heuer and Rubner, 2012). The states of nature dictated by football match outcomes matter significantly, economically or otherwise; teams progress in competitions, fans gain bragging rights and joy, and bettors may make returns (or losses). While the result generally determines the state of nature that matters to these different agents (e.g., winning a round-robin or knock-out competition), the scoreline can sometimes be the first tie-breaker after the result. League positions and championships, when teams are tied on cumulative points totals from results, are frequently determined by some function of scorelines (e.g., the difference between goals scored and conceded, or head-to-head records between teams over multiple matches). Some cup competitions have scoreline-related tie-breaker rules, such as 'away goals'.[1] Fundamentally though, the result is a function of the scoreline.

The majority of attention in the academic literature on forecasting football has focused on results, rather than scorelines, perhaps due to the more complicated nature of the latter (e.g., Angelini and De Angelis, 2019; Forrest and Simmons, 2000; Forrest et al., 2005; Goddard, 2005). But scorelines also matter. Based on our observation and a rough estimation from the world's largest sports betting exchange in 2019, *Betfair Exchange*, the exact scoreline in a football match is a popular outcome to predict and bet on: focusing on pre-kick-off markets for several important matches (i.e., high liquidity markets, of £1million or more matched bets, e.g., the English Premier League or competitive internationals), for every £1.00 worth of bets made on the result outcomes of a match, approximately £0.20 worth of bets are made on the exact scoreline markets in the same match. This compares with £0.70 worth of bets placed on the total number of goals scored in a match, £0.25 on the Asian Handicap markets, and £0.20 on the margin (goal difference) between the two teams at the end of a match. Notably, these other mentioned match outcomes and popular prediction markets are all functions of the final scoreline. As there are only three possible outcomes for the result, and many times more potential outcomes for the scoreline, it follows that forecasting the scoreline is more difficult. Historically, the most likely result outcome from a football match is a home win (occurring roughly 48% of the time), while the most likely scoreline outcome is a 1-1 draw (occurring roughly 11% of the time).[2]

[1] For example, in the UEFA Champions League, if two teams are equally matched after playing each other twice, home and away, i.e., the cumulative scoreline is a draw, then the team which has scored more goals away from home is the winner.

[2] Author calculations using the entire history of football matches listed on *Soccerbase.com*, i.e., from 511,759 recorded matches up to 8 January, 2019.

Within economic forecasting in recent decades there has been a trend toward probability (or density) forecasts: attaching probabilities to the different possible outcomes of an event or time series. Bookmakers essentially produce density forecasts by offering odds on a range of different scorelines. Well-established statistical methods for predicting scorelines also generate probability forecasts. In this study, we evaluate a standard statistical model of football match scoreline forecasting, assuming a bivariate Poisson distribution for goals, which follows from various previous contributions to the football forecasting literature (e.g., Maher, 1982; Dixon and Coles, 1997; Goddard, 2005; Karlis and Ntzoufras, 2005). In particular, we compare the model's performance over two seasons of the English Premier League (EPL), 2016/17 and 2017/18, against betting markets. Therefore, we also treat the betting markets (or odds setters) as probability forecasters.

We evaluate these two sets of forecasts primarily using the Mincer and Zarnowitz (1969) regression-based framework. We find that both bookmakers and the statistical model appear to be biased in terms of predicting scoreline outcomes. However, forecasts from neither source are generally biased for result outcomes. We also carry out an analysis of forecast encompassing (e.g., Chong and Hendry, 1986; Fair and Shiller, 1989). The statistical model does encompass the scoreline odds-implied forecasts from betting markets. In other words, the model probabilities provide 'better' forecasts of the football match scorelines over the two seasons studied. However, this is not sufficient that a simple betting strategy based on the model forecasts for scoreline outcomes would have generated a positive return on investment, using averages of the odds available, not least because these odds implied a particularly high profit margin (or overround) for bookmakers of approximately 12%. We also find no evidence that this simple betting strategy, based on the model forecasts, would have generated positive returns on the markets for either the margin of victory or the total goals scored in a match. However, there is some evidence that the model would have generated marginally positive returns when betting on result outcomes.

Several papers have previously looked at odds setters as football match forecasters. Forrest et al. (2005) studied bookmakers in the 1990s and 2000s, finding that they were increasingly accurate during this time, reflecting growing commercial pressure in the industry. Štrumbelj and Šikonja (2010) updated this finding, but highlighted an aspect of the strangeness of football match scorelines: the draw. These authors found that bookmaker odds provided little predictive information on the relative frequency of draws, and noted that Pope and Peel (1989) and Dixon and Pope (2004) had found something similar in earlier studies. Štrumbelj and Šikonja (2010) suggested that this reflected the residual nature of the draw outcome; it is the remaining probability mass after the home and away teams' strengths have been accounted for. Angelini and De Angelis (2019) studied the odds of online bookmakers on football matches in 11 top European professional leagues between 2006 and 2017. Using a forecast-based approach, they tested whether these markets were generally efficient, finding that they were in most countries, even if the best odds on match outcomes were selected from among bookmakers. This result was further supported by Elaad et al. (2019), who found that after accounting for heterogeneity among online bookmakers, the prices set on result outcomes in the EPL and the rest of professional English football were generally unbiased as forecasts. However, Angelini et al. (2019) have found significant evidence of mispricing and bias in a betting exchange, specifically *Betfair Exchange*, for the EPL, both in pre-match and in-play odds. Dixon and Pope (2004) is one of the contributions in the literature that has considered football scoreline outcomes rather than just results, finding that the markets for exact scoreline predictions were generally inefficient in the 1990s. To some extent, we are updating their study here.

There is a substantial literature studying the behavioural biases implied by sports forecasts, not least betting odds. Perhaps most famously and extensively studied is the favourite-longshot bias, whereby the probability forecasts implied by prediction market prices typically suggest that favourites, i.e., those most likely to win, are underbet. Rational explanations of this bias focus on the potential for relative risk-love among gamblers (see for a summary Ottaviani and Sørensen, 2008). In football, Cain et al. (2000) showed that this bias appears in UK football results odds, though Angelini and De Angelis (2019) found less convincing evidence in more recent years throughout the European betting market odds for match results. We find evidence here of significant favourite-longshot bias in football match scoreline odds, though none for result outcomes. In other words, the betting markets would appear to overestimate the likelihood of a rare scoreline, such as 4-4, over more common ones, such as 1-0 or 1-1. This may be consistent with behavioural or misperception-based explanations of the favourite-longshot bias, such as bettors not being able to distinguish between events with different low probabilities of occurring (e.g., Snowberg and Wolfers, 2010).

There are many previous studies statistically modelling the outcomes of football matches, and which have subsequently evaluated the forecasting performance of such models against betting markets, mostly focusing on result outcomes. Maher (1982) analysed both the independent and the bivariate Poisson processes of goal arrival, while Dixon and Coles (1997) adjusted that model to account for a tendency toward low-scoring and close matches, a common feature of English football in the early 1990s, the period they were focused on. Like ourselves, Dixon and Coles (1997) were interested in the potential for inefficiencies in betting markets, considering whether betting on home or away wins based on their model forecasts could generate consistently positive returns. Boshnakov et al. (2017) introduced a bivariate Weibull count model of goals to this topic, which they documented as improving upon the Poisson model of Dixon and Coles (1997) or Karlis and Ntzoufras (2005). Like Dixon and Coles (1997), they evaluated their model's forecasts by using it to inform a potentially successful betting strategy, looking at both result outcomes and whether more than 2.5 goals were scored in a match. Similarly, Buraimo et al. (2013) have demonstrated that betting whenever positive returns were expected based on the University of Warwick's 'Fink Tank' statistical model's probability forecasts, which were published in a British newspaper, could have generated positive expected returns on match result outcomes for each EPL season between 2006/07 and 2011/12.

The rest of the article is organised as follows: Section 2 introduces the data; Section 3 sets out the methodology we employ; Section 4 presents our results; and Section 5 concludes.

2 Data

The EPL is generally regarded as the foremost domestic club competition globally.[3] For our sample of forecasts, we consider the 380 matches played in each of the 2016/17 and 2017/18 EPL seasons. We focus on these two recent seasons so that our results are relevant to how sports betting markets function today, given the rapid change to this industry sector over the past few decades, not least due to the increased competition, as most of the activity has moved online and away from the high street (Forrest, 2008). We extract data on the outcomes of football matches from *Soccerbase.com*, including for the EPL seasons before 2016/17 to estimate the statistical forecasting model, which is described later.

The right panel of Table 1 displays the distribution of match results in the 2016/17 and 2017/18 EPL seasons, showing that there were more home wins in 2016/17 than in the following season, and fewer draws. As home wins happen almost half the time, this provides a naïve forecasting method. Forrest and Simmons (2000) documented that newspaper tipsters tended to have a lower success rate than such a naïve forecasting method as always picking the home team to win. Table 2 presents the distribution of scorelines across the two seasons that we focus on. The left panel is the 2016/17 season and the right panel is the 2017/18 season. There were 33 different unique scorelines in 2016/17 and 32 in 2017/18, of which around two thirds involved each team scoring at most two goals. Within each panel, the rows represent the number of goals scored by the home team, and the columns give scorelines where the away team scored a particular number of goals. The top left entry in each panel is a 0-0 draw. 7.1% of matches in 2016/17 and 8.4% in 2017/18 had 0-0 scorelines. There were slightly more draws in 2017/18 than 2016/17, and fewer goals, but these differences between the two seasons are generally not statistically significant.

2.1 Bookmaker odds

While bookmakers exist to profit maximise rather than forecast event outcomes per se, to do the former they must do the latter sufficiently well. We consider the decimal odds d set by a bookmaker. Decimal odds are inclusive of the stake (the money amount bet), such that if the potential event outcome being bet on occurs, the bettor is paid dz, where z is the stake. If it does not occur, then the bettor loses their stake z. The implied outcome probability of a given decimal odd set is $p = 1/d$. Decimal odds relate to the traditional UK presentation

[3]It is a derivative of the Football League, founded in 1888. The total club revenues for the EPL at £5.3bn are almost equal to the sum of the next two leagues combined, Spain's La Liga (£2.9nb) and Germany's Bundesliga (£2.8nb) (see 2018 Deloitte Annual Review of Football Finance; www2.deloitte.com/uk/.

of fractional odds, f, by $d = f + 1$. In reality, there is an overround (or vig) included in the prices of the outcome set for any given event and bookmaker; if the implied probabilities for all outcomes in the event space are summed, then they will add to more than one. Various methods have been suggested to correct for the overround such that applied researchers can then interpret posted bookmaker odds as implied probabilities (see the summary of these methods by Štrumbelj, 2014, as well as Manski, 2006, for a theoretical discussion on interpreting betting prices as implied probabilities). In the analysis which follows, we use the most simple of these corrections, by dividing the raw probability implied by the quoted odds on each outcome by the booksum, which is the sum of all the implied quoted probabilities offered for the various possible outcomes on some event (e.g., over all possible scorelines offered).[4]

We obtain the bookmaker odds for all EPL match outcomes listed on *Oddsportal.com*, where in this study we will use the odds for the result, scoreline, margin of victory (plus/minus x goals) and the total number of goals scored. From this source, we have information from 51 individual bookmakers, and also a betting exchange, *Matchbook*. The odds reflect what was offered immediately before matches kicked off. The left panel of Table 1 presents the average among these sources of the odds-implied probability for the three different match result outcomes, without adjusting for the overround. Betting market prices were more consistent in the period we study than the actual match outcomes, predicting in both seasons the home teams to win 46% of the time, the away teams to win 32% of the time, and the draw to occur 25% of the time (implying an overround of about 4%). In the right panel of Table 1, we present the actual frequencies, suggesting that bookmakers tended to over-estimate the likelihood of an away win. Table 3 presents the implied probability, or frequency, from the average bookmaker odds for each match scoreline in each season. To demonstrate how diverse these predictions are, in 2016/17, at least some bookmakers offered odds on scorelines of 7-4, 7-5, 7-6 and 6-7 for the Premier League, but in 2017/18 such odds were never offered. In the entire history of the (English) Football League since 1888, of more than 220,000 matches, there have been twenty-one 7-4 scorelines, five 7-5 scorelines, and no 7-6 or 6-7 scorelines. The scoreline odds-implied probabilities indicate a sizeable average overround of about 12%, with the majority of implied probabilities being higher than the actual proportions from Table 2. This compares with an average overround of about 4% for the result outcomes. As for the result outcomes, variation between the two seasons in the odds-implied scoreline frequencies is smaller than in the actual proportions of scoreline outcomes.

3 Methodology

To compare and evaluate the implied bookmaker forecasts described above, we generate a set of probability forecasts using a statistical model. The model we select for this purpose

[4]The implied probability of match outcome i from the bookmaker odds is then given by: $p_i = (1/d_i)/\sum_i(1/d_i)$.

is well-known, and could be arguably considered as the 'standard' or 'benchmark' statistical model for football match scoreline forecasting (e.g., Goddard, 2005). We briefly describe our application of this model in Section 3.1. In Section 3.2, we discuss the methods we will use to evaluate and compare both the bookmakers as forecasters and the statistical model.

3.1 Scoreline forecasts from a 'standard' statistical model

To create scoreline forecasts, we first estimate the goal arrival process in football matches using a bivariate Poisson regression model, of the form proposed (and coded) by Karlis and Ntzoufras (2003, 2005), which is also based on the original Maher (1982) and Dixon and Coles (1997) approaches, and which is applied by Dixon and Pope (2004); i.e., the goals scored by each team in a football match are modelled as jointly Poisson distributed. The counts of goals scored in match i for the home and visiting teams can be thought of as functions of their own strengths, X_{i1} and X_{i2}, respectively, and some third common factor X_{i3}, representing the match conditions (e.g., weather, time of the year). If the goals of the home team in match i are denoted by h_i, and those of the visiting team by a_i, then we can define three Poisson distributed random variables X_{i1}, X_{i2}, X_{i3}, such that $h_i = X_{i1} + X_{i3}$ and $a_i = X_{i2} + X_{i3}$. We assume that these are jointly distributed according to a bivariate Poisson distributed, with $BP(\lambda_{i1}, \lambda_{i2}, \lambda_{i3})$. The regression model can be written as:

$$\begin{aligned} (h_i, a_i) &\sim BP(\lambda_{i1}, \lambda_{i2}, \lambda_{i3}) \ , \\ \log(\lambda_{ik}) &= \mathbf{w}'_{ik}\boldsymbol{\beta}_k \ , \qquad k = 1, 2, 3 \ , \end{aligned} \tag{1}$$

where $\mathbf{w}_{ik}$ is a vector of explanatory variables, and $\boldsymbol{\beta}_k$ is a potentially large vector of coefficients, to be estimated along with the λ parameters. We include fixed effects for both teams in a match in $\mathbf{w}_{ik}$, for each $k = 1, 2$, to allow for teams having particular goal scoring or defending strengths irrespective of who their opponent is. The explanatory variables also include day of the week and month fixed effects for the modelling of λ_{i3}, to reflect the fact that midweek matches may have different properties to weekend ones, and matches in the middle of winter may be different to those in the autumn or spring. We also add an indicator in the λ_{i3} equation for whether a match follows a break in the season for international matches. We include information in the model about the varying lagged league positions and the recent form of each team, following the application in Goddard (2005). We also include our calculations of each team's measured Elo (1978) strengths as they varied throughout the season, based on the historical results for all relevant teams, including those not playing in the EPL in the period studied. The Elo rankings, and the predictions that they imply for match outcome probabilities, are commonly used to estimate the relative strengths of football teams, both in practical applications (e.g., https://www.eloratings.net/) and in academic research (e.g., Hvattum and Arntzen, 2010). We also add variables to the λ_{i1} and λ_{i2} equations for whether each team is still in the main domestic cup competition, the FA Cup, at the time of the current EPL match, i.e., whether they have already been knocked out. Goddard (2005) found this to matter for goal arrival in league matches, and others have found this to matter for league attendance, and attendance to matter for home advantage. We also add variables to these equations for whether each team can still achieve a top-two position in the league, and a variable for whether a team is returning to domestic action having played in European competition in their previous match, since this may affect squad rotation and player tiredness, and thus goal scoring or defending.

The statistical model is estimated by maximum likelihood up to each round of ten matches between the twenty EPL teams in each season, using the past calendar year of matches, and the estimated parameters are subsequently used to make predictions. The values of $\widehat{\lambda}_{ik}$ are used to generate probabilities for a range of scorelines of the upcoming match. By summing over these scorelines, probability forecasts of the three different result outcomes can also be generated. Combinations of the λs give predictions of the mean (or expected) number of goals scored within matches, as well as the Poisson goal scoring rate of each team.

To test the efficiency of the bookmaker markets in the 2016/17 and 2017/18 EPL seasons, using a simple betting strategy, we generate scoreline point forecasts (picks) in two ways. First, we use whatever the statistical model outputs as the most likely scoreline as the pick, which we call *Unconditional* forecasts. Second, we condition the scoreline pick on the most likely forecast result outcome. In this case, if all the probabilities of the home win scorelines sum to a larger number than all the probabilities of either the draw or the away win scorelines, then we would choose the most likely home win scoreline as the pick. We call these *Conditional* forecasts; i.e., conditional on the most likely result outcome, what is the most likely scoreline? This tends to generate differences, as empirically the most common scoreline is a 1-1 draw (see Table 2), but the most likely result outcome is a home win.

3.2 Forecast evaluation and comparison

The issue of forecasting football match scorelines is interesting along a number of dimensions. In particular, the difficulty of the task is emphasised by considering the variation in goals scored by teams over matches. In our forecast sample of 760 EPL matches over the two seasons, the mean number of goals scored per match is 2.73 and the variance is 2.78. Conditional on a home win, the variance of home goals is 1.5 and the variance of total goals is 2.7, while conditional on an away win occurring, the variance of away goals is 1.3 and the variance of total goals is 2.3. Furthermore, any match has a number of outcomes and sub-outcomes that can matter in terms of how scoreline forecasts are evaluated. Each of the following main outcomes could be considered when asking whether bookmaker odds reflect accurate or efficient forecasts of match scoreline outcomes:

Scoreline: the actual numbers of goals scored by each team in match i. The scoreline is a pair of numbers, $\mathbf{s}_i = (h_i, a_i)$, where the number of goals scored by the home team is always listed first. Throughout what follows, we denote the actual scoreline by $\mathbf{s}_i$ and any forecast of it by $\widehat{\mathbf{s}}_i$, etc.

Result: whether either team wins the match, or it ends in a draw. We denote the result of some match i as r_i. The result can be defined as a single variable taking three values: one each for a home win, an away win and a draw. For example, we could define the following values:

$$r_i = r(\mathbf{s}_i) = \begin{cases} 0 & \text{if } h_i < a_i \\ 0.5 & \text{if } h_i = a_i \\ 1 & \text{if } h_i > a_i \ . \end{cases} \tag{2}$$

Note that the result r_i is a function of the scoreline, so $r_i = r(\mathbf{s}_i)$.

Margin: the difference between the goals scored by the two teams in match i; $m_i = m(\mathbf{s}_i) = h_i - a_i$.

Total goals scored: the total number of goals scored by both teams in match i; $t_i = t(\mathbf{s}_i) = h_i + a_i$.

3.2.1 Return on investment

Evaluating scoreline forecasts according to betting prices is arguably the most natural evaluation method, since it reflects the potential payoffs from making decisions based on those forecasts. It can also tell us whether these markets are efficient, in so far as whether the readily-available information and methods used by our statistical forecasting model are already reflected in market prices. If not, and the model generated forecasts imply a consistently profitable betting strategy, then these markets might be determined as inefficient. In the case of the bookmaker exact scoreline markets, the average overround is relatively high at 12%, as discussed above. In which case, the statistical forecasting model would need to be substantially more accurate than the odds-implied predictions of the bookmakers for there to be any simple profitable strategy based on the former.

We calculate the returns from betting on the result, scoreline, margin or total goals scored in a match, otherwise referred to as a return on investment (ROI), as follows. If d_i are the decimal odds in match i for the scoreline (or other outcome) consistent with the forecast $\widehat{s}_i$, then the ROI from a one unit bet on that event outcome would be:

$$ROI_i = d_i \mathbb{1}\{\mathbf{s}_i = \widehat{\mathbf{s}}_i\} - 1 \ . \tag{3}$$

Throughout our analysis, for scorelines and the over-under markets of total goals scored or the margin of victory, we use the mean of the bookmaker odds that we collected. In the case of results, we take the best available bookmaker odds among those collected, all as posted right before matches began.

3.2.2 Regression-based methods and forecast encompassing

The following is based on the Mincer and Zarnowitz (1969) regression-based forecast evaluation framework and extensions thereof. If we denote $\widehat{p}_{ij}$ as our probability forecast of match i for event outcome j, and y_{ij} as the relevant binary specific outcome (e.g., a scoreline), taking a value of one if that outcome happened and zero if not, then the linear regression model is given by:

$$y_{ij} = \alpha + \beta \widehat{p}_{ij} + \epsilon_{ij} \ , \tag{4}$$

where α and β are the intercept and slope coefficients, respectively, and ϵ_{ij} is the the error term. The weak efficiency of a forecast depends on the restriction $\alpha = 1 - \beta = 0$ holding. A stronger test of efficiency includes other information available at the forecast origin, and can be tested using the regression model:

$$y_{ij} = \alpha + \beta \widehat{p}_{ij} + \mathbf{z}_i' \boldsymbol{\gamma} + \nu_{ij} \ , \tag{5}$$

where $\mathbf{z}_i$ is a vector of potentially other important variables for explaining the outcome, y_{ij}, and ν_{ij} is the error term. Strong efficiency further requires that $\boldsymbol{\gamma} = \mathbf{0}$ holds in addition. If $\boldsymbol{\gamma} \neq \mathbf{0}$, then other known information at the forecast origin is relevant and the forecast is not efficient.

Taking expectations of (4) yields that for unbiasedness we require $E(\widehat{p}_{ij}) = \alpha/(1-\beta)$. To test for this, we could estimate the regression:

$$\widehat{e}_{ij} = \theta + \nu_{ij} \ , \tag{6}$$

where $\widehat{e}_{ij} = y_{ij} - \widehat{p}_{ij}$ is the forecast error and ν_{ij} is the error term, with the null hypothesis that $\theta = 0$. Strictly speaking, in addition to the hypothesised restrictions holding, we require that the residuals from each regression estimation are approximately normally distributed, and free from any autocorrelation or heteroskedasticity. In their application for newspaper tipsters' football match forecasts, Forrest and Simmons (2000) add a range of variables that are public information into $\mathbf{z}_i$, including the recent results of each team and league-standing-related information. We do similarly, by using our derived dynamic Elo (1978) ratings of teams, and the implied predicted match outcome probabilities from these ratings. When testing the efficiency of the scoreline forecasts, we also include in $\mathbf{z}_i$ the historic frequency of each scoreline, the current league points of the home team, the recent form of the home team, measured by the number of league points gained in the their last six matches, and for the latter two variables we also include the difference between the home and away teams in these values.

Other forecasts could be added to this regression analysis. In doing so, we could test whether any of the various forecasts are *encompassing* one another. A forecast a is said to encompass forecast b if it can explain variation in the forecast errors from forecast b, and forecast b cannot explain any of the variation in the forecast errors from forecast a:

$$\widehat{e}_{ija} = \theta_a + \phi_a \widehat{p}_{ijb} + \nu_{ija} \ , \tag{7}$$

$$\widehat{e}_{ijb} = \theta_b + \phi_b \widehat{p}_{ija} + \nu_{ijb} \ , \tag{8}$$

and $\mathbf{H_0}: \ \phi_a = 0, \phi_b \neq 0$, i.e., can one forecast explain what another forecast cannot? Chong and Hendry (1986) and Fair and Shiller (1989) both consider the possibility of encompassing in this manner. If $\phi_a \neq 0$ *and* $\phi_b \neq 0$, then a linear combination of the forecasts would be on average more effective than taking any single forecast in isolation. For example, focusing on the case of the bookmaker implied probabilities, in this way we can test whether our generated statistical model probabilities add any information when trying to determine the accurate probability of a future football match outcome taking place. If we find that the statistical model forecasts do encompass the bookmaker implied ones, then we could conclude quite simply that the former are *better* forecasts.

4 Results

4.1 Forecast efficiency

In this section, to evaluate individually and comparatively the statistical model and betting markets as sources of football match forecasts, we describe the results of Mincer and Zarnowitz (1969) regression-based efficiency tests. We pool the 2016/17 and 2017/18 EPL seasons, so the number of match forecasts studied in each of these regressions is 760. When we refer to Model forecasts, we are evaluating the probability forecasts produced using the bivariate Poisson model set out in Section 3.1. By Bookmaker forecasts, we are referring to the implied probabilities of outcomes derived from odds, as described before.

Table 4 presents the outcomes from regressions evaluating the weak and strong efficiency of scoreline forecasts as per Equation (4) and Equation (5), respectively, with a column for each forecast type. Across both forecast methods in the strong efficiency cases (columns (3) and (4), Table 4), the additional variables in the regressions are insignificant, i.e., $\boldsymbol{\gamma}$ in Equation (5) is insignificant from $\mathbf{0}$. This means that the weak efficiency testing results (columns (1) and (2), Table 4) are practically identical. This is not unexpected. While these team-specific variables must matter for result outcomes, given the sheer number of possible scoreline outcomes they simply are not important. It might be anticipated that the historical frequency of each scoreline would be significant, but our findings suggest that this is factored into each type of forecast. The bottom row of Table 4 reports an F-test of strong efficiency, which here is the null hypothesis that $\alpha = 0$, $\beta = 1$, and $\boldsymbol{\gamma} = \mathbf{0}$. The null hypothesis is heavily rejected in each case at standard levels of significance. In other words, the forecasts are suggestively not efficient and there is evidence of mispricing in the betting markets. The $\widehat{\beta}$ coefficient on the Bookmaker forecasts, 1.16, is significantly greater than one at standard levels, which is indicative of the well-known favourite-longshot bias. Hence, we can document the existence of this bias among football match scorelines odds, whereas it has typically only been described for result outcomes in the previous literature. This implies that a profitable betting strategy for scorelines, if it exists given the magnitude of the overround in these markets, is likely to be attained by betting on favourites (short odds, e.g., 1-1) more frequently than on longshots (long odds, e.g., 4-4). In contrast, the Model provides forecasts in these two seasons which exhibit a significant *reverse* favourite-longshot bias for scorelines. This suggests that the Model, and perhaps the assumed Poisson distribution of the goal scoring in football, is biased against high scoring matches.

We also consider the (implied) probability forecasts of the three different match result outcomes. In Table 5, we present the weak and strong efficiency regression test results, estimating equivalent regression models as before with scorelines, i.e., Equation (4) and Equation (5), including the Elo-ranking based predicted match outcome as an explanatory variable. For the draw outcome, we take the squared difference of the Elo prediction from 0.5, referring to this as a 'Balance' measure.[5] The table of results has three panels: the top panel for the home win outcome, the middle panel for the draw, and the bottom panel for the away win. We also present the F-test of efficiency (null hypothesis of $\alpha = 0$, $\beta = 1$ and $\boldsymbol{\gamma} = \mathbf{0}$). Despite some individually significant coefficients for γs, the test nonetheless does not reject the null of strong efficiency for the Model and Bookmaker forecasts in all three outcome cases at standard levels. The results are qualitatively the same for the weak efficiency tests. The $\widehat{\beta}$ coefficient on the Bookmaker forecasts is only significantly different from one for the away win at standard levels, when including the ELO prediction as an extra explanatory variable. This suggests that the typical favourite-longshot bias for football match results only shows up in the away win odds in the EPL during this period and on average amongst the considered sample of bookmakers. As for the scorelines, the Model shows evidence of generating forecasts which exhibit a significant *reverse* favourite-longshot bias, implying that it too infrequently predicts surprising match outcomes.

4.1.1 Forecast encompassing

We now consider the outcomes of encompassing regressions, described by Equations (7)-(8). We apply the bilateral regression encompassing tests for the Model and Bookmaker probability forecasts over all 760 sample matches and for all scorelines which bookmakers posted odds on. The forecast encompassing results are summarised in Table 6. This shows the t-statistics for the equivalent of the estimated ϕ_a and ϕ_b coefficients. The results are presented such that the row is the particular forecast error in the regression equation (the dependent variable), and the column is the other forecast being added into the model (the explanatory variable). Hence for the Model probabilities, the entry in the first row and column is blank, since we cannot enter the Model probability forecast into the Model probability forecast error regression model. We highlight t-statistics that are very significant, i.e., 3.8 or larger, based on the rule of thumb established in Campos et al. (2003) for adjusting

[5] As the Elo prediction lies on the unit interval, where 0 implies a certain away win and 1 a certain home win, we can take 0.5 to imply a 'certain' draw.

t-statistics with large sample sizes (here it is 61,560). Using our notation and definition of encompassing from before, reading from right to left in the table for a particular source of forecast errors a (b), the t-statistics give the values of ϕ_b (ϕ_a) for the other source of forecasts (column). When asking if the Model probabilities (a) encompass the Bookmakers (b), $\{\widehat{\phi}_b : t\text{-stat} = 1.80\}$ and $\{\widehat{\phi}_a : t\text{-stat} = \mathbf{8.77}\}$. To repeat, one forecast source is said to encompass another if $\mathbf{H_0} : \phi_a = 0, \phi_b \neq 0$, and vice versa if $\mathbf{H_0} : \phi_a \neq 0, \phi_b = 0$. If $\phi_a \neq 0$ *and* $\phi_b \neq 0$, then a linear combination of such forecasts would be more effective than taking any single forecast in isolation. The results do show that the Model probabilities (a) significantly encompass the Bookmaker probability forecasts. Therefore, we can conclude that the 'standard' statistical model for forecasting football match scorelines dominates the Bookmaker odds-implied forecasts, and that it is in a sense *better* at this prediction job. This is consistent with other attempts in the literature to compare statistical models and bookmakers as football match forecasters (e.g., Dixon and Pope, 2004; Buraimo et al., 2013; Boshnakov et al., 2017), though in these previous cases the comparisons used betting strategies and returns on investment, and, apart from Dixon and Pope (2004), they focused on match results rather than scorelines.

4.2 A simple betting strategy

Table 7 shows the returns on investment from systematically betting before every match the same amount on the outcomes implied by the point forecasts from the statistical model. In other words, these returns are derived by assuming that the forecaster used their scoreline point forecast, for each of the 380 matches in a season, to place a £x bet on each of the markets for the correct result, the correct scoreline, the margin being equal or greater than that implied by the predicted scoreline, and the total goals being equal or greater than that implied by the predicted scoreline. We consider two sets of point forecasts derived from the statistical model, Unconditional and Conditional, as defined earlier.

In general, betting on results based on the statistical model could have generated positive returns. However, this assumes that the bettor made use of the best available odds from the range of bookmakers available in the UK. Over the 2016/17 season, a ROI of 12.7% was possible using this simplest of betting strategies for result outcomes, if following the Conditional pick from the model. Surprisingly, the Unconditional picks provided a positive 4.8% ROI over the 2017/18 season, whereas the the Conditional picks provided a ROI of -0.2%, despite the latter reflecting the most likely result outcomes according to the model and the former not doing so.

Over the 2016/17 season, the model point forecasts provided negative ROIs from betting using the average scoreline odds in the sample of 51 bookmakers, though these were smaller in magnitude than the average overround of 12%. However in 2017/18, both sets of model picks would have implied substantially more negative ROIs, with losses of more than 25%. Despite the efficiency testing results of Section 4.1 demonstrating that the average bookmaker correct scoreline odds appear to be mispriced, with a favourite-longshot bias, and the encompassing regression results suggesting that the statistical model is a better forecaster, the simple betting strategy is generally not successful. In other words, without devising a more complicated strategy, such as identifying 'value' bets, there is no significant evidence from the statistical model suggesting that the UK betting markets for football match outcomes are inefficient, based on common methods and using readily available historical information about match features and outcomes. In part, this could be an indictment of the 'benchmark' statistical model, which we have earlier showed tends to significantly under-predict the frequency of high scoring matches.

To put these returns in perspective, we also consider what the bettor would have earned from systematically betting the same amount on the home win in every match. As mentioned previously, this strategy may be naïve, but it has been shown to outperform semi-expert (newspaper) tipsters in the past in English football (Forrest and Simmons, 2000). The ROI over the 2016/17 and 2017/18 seasons using the best result outcome odds from this strategy would have been 9.8%. The ROIs from two similarly naïve strategies based on average scoreline odds are -21.9% and -12.6% from always betting on 1-1 and 1-0, respectively,over the same period.

5 Conclusion

We have studied the forecasts of scorelines in association football matches. We compared the odds-implied probability forecasts of bookmakers against those we generated ourselves from a standard statistical model. We found that over two seasons of EPL matches, 2016/17 and 2017/18, both sources of forecasts were generally inefficient for exact scoreline outcomes. The model-based forecasts tended to under-predict high-scoring and less likely outcomes, whereas the bookmaker forecasts implied an over-prediction of unlikely scorelines. In spite of this, both sets of forecasts were efficient at predicting match result outcomes. There was some evidence that the scoreline model was 'better' than the bookmakers. This difference was not enough that a simple and systematic betting strategy, based on point forecasts from the model, could generate positive financial returns. However, the evidence of significant mispricing in scoreline odds, despite the large overround set by bookmakers, does suggest that an alternative statistical model could in theory generate greater financial returns on football match scorelines than result markets, especially if it applied the odds from betting exchanges.

There is substantial room for further research in this area. It remains the case that the wider literature in this area, which studies either the practice of forecasting or issues of financial market efficiency, has not paid much regard to the diverse range of betting and prediction market data available for any given event. For example, we know of no study which has studied how the prices, liquidity and volumes on different markets for the same event on betting exchanges co-move, or whether the way in which they move together (or not) reveals sizeable inefficiencies, or whether any of this suggests particular behavioural responses to the arrival of new information.

References

Angelini, G., L. D. Angelis, and C. Singleton. 2019. "Informational efficiency and price reaction within in-play prediction markets." Economics & Management Discussion Papers, Henley Business School, Reading University.

Angelini, G., and L. De Angelis. 2019. "Efficiency of online football betting markets." *International Journal of Forecasting*, 35(2): 712–721.

Boshnakov, G., T. Kharrat, and I. McHale. 2017. "A bivariate Weibull count model for forecasting association football scores." *International Journal of Forecasting*, 33(2): 458–466.

Buraimo, B., D. Peel, and R. Simmons. 2013. "Systematic Positive Expected Returns in the UK Fixed Odds Betting Market: An Analysis of the Fink Tank Predictions." *International Journal of Financial Studies*, 1(4): 1–15.

Cain, M., D. Law, and D. Peel. 2000. "The Favourite-Longshot Bias and Market Efficiency in UK Football Betting." *Scottish Journal of Political Economy*, 47(1): 25–36.

Campos, J., D. Hendry, and H.-M. Krolzig. 2003. "Consistent Model Selection by an Automatic Gets Approach." *Oxford Bulletin of Economics and Statistics*, 65(s1): 803–819.

Chong, Y. Y., and D. F. Hendry. 1986. "Econometric evaluation of linear macro-economic models." *The Review of Economic Studies*, 53(4): 671–690.

Dixon, M. J., and S. C. Coles. 1997. "Modelling association football scores and inefficiencies in the football betting market." *Applied Statistics*, 47(3): 265—280.

Dixon, M., and P. Pope. 2004. "The value of statistical forecasts in the UK association football betting market." *International Journal of Forecasting*, 20(4): 697–711.

Elaad, G., J. J. Reade, and C. Singleton. 2019. "Information, prices and efficiency in an online betting market." *Finance Research Letters*, In Press.

Elo, A. E. 1978. *The rating of chessplayers, past and present.* London Batsford.

Fair, R. C., and R. J. Shiller. 1989. "The Informational Context of Ex Ante Forecasts." *The Review of Economics and Statistics*, 71(2): 325–331.

Forrest, D. 2008. "Soccer Betting in Britain." In *Handbook of Sports and Lottery Markets.* Eds. by D. B. Hausch, and W. T. Ziemba, San Diego Elsevier, 421 – 446.

Forrest, D., J. Goddard, and R. Simmons. 2005. "Odds-Setters As Forecasters: The Case of English Football." *International Journal of Forecasting*, 21(3): 551–564.

Forrest, D., and R. Simmons. 2000. "Forecasting Sport: The Behaviour and Performance of Football Tipsters." *International Journal of Forecasting*, 16(3): 317–331.

Goddard, J. 2005. "Regression Models for Forecasting Goals and Match Results in Association Football." *International Journal of Forecasting*, 21(2): 331–340.

Heuer, A., and O. Rubner. 2012. "How Does the Past of a Soccer Match Influence Its Future? Concepts and Statistical Analysis." *PLOS One*, 7(11): 1–7.

Hvattum, L. M., and H. Arntzen. 2010. “Using ELO ratings for match result prediction in association football.” *International Journal of Forecasting*, 26(3): 460–470.

Karlis, D., and I. Ntzoufras. 2003. “Analysis of Sports Data Using Bivariate Poisson Models.” *Journal of the Royal Statistical Society (Statistician)*, 52(3): 381–393.

Karlis, D., and I. Ntzoufras. 2005. “Bivariate Poisson and Diagonal Inflated Bivariate Poisson Regression Models in R.” *Journal of Statistical Software*, 14(10): .

Maher, M. J. 1982. “Modelling association football scores.” *Statist. Neerland.*, 36(3): 109–118.

Manski, C. 2006. “Interpreting the predictions of prediction markets.” *Economic Letters*, 91(3): 425–429.

Mincer, J., and V. Zarnowitz. 1969. “The evaluation of economic forecasts.” In *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*. NBER, 1–46.

Ottaviani, M., and P. N. Sørensen. 2008. “The Favorite-Longshot Bias: An Overview of the Main Explanations.” In *Handbook of Sports and Lottery Markets*. Eds. by D. B. Hausch, and W. T. Ziemba, San Diego Elsevier, 83–101.

Pope, P. F., and D. A. Peel. 1989. “Information, Prices and Efficiency in a Fixed-Odds Betting Market.” *Economica*, 56(223): 323–341.

Snowberg, E., and J. Wolfers. 2010. “Explaining the Favorite-Longshot Bias: Is It Risk-Love or Misperceptions?” *Journal of Political Economy*, 118(4): 723–746.

Štrumbelj, E. 2014. “On determining probability forecasts from betting odds.” *International Journal of Forecasting*, 30(4): 934–943.

Štrumbelj, E., and M. Šikonja. 2010. “Online bookmakers’ odds as forecasts: The case of European soccer leagues.” *International Journal of Forecasting*, 26(3): 482–488.

TABLE 1: Result outcomes in the 2016–17 and 2017–18 EPL seasons (%): comparison of actual outcomes with the average implied frequency from bookmaker prices

	Bookmakers			Actual		
Season	Home	Draw	Away	Home	Draw	Away
2016/17	46.1	25.3	32.3	49.2	22.1	28.7
2017/18	46.3	25.3	32.4	45.5	26.1	28.4

Source: author calculations using *Oddsportal.com* and *Soccerbase.com*

TABLE 2: Frequency of scoreline outcomes in the 2016–17 and 2017–18 EPL seasons (%).

		2016–17 Away goals								2017–18 Away goals						
		0	1	2	3	4	5	6	7	0	1	2	3	4	5	6
	0	7.1	5.5	4.5	2.6	1.8	0.3	0.0	0.0	8.4	6.1	3.9	3.2	1.8	0.0	0.3
	1	10.0	10.0	6.3	3.2	1.8	0.3	0.3	0.3	11.6	11.8	6.3	1.3	1.8	0.3	0.0
	2	8.7	7.9	4.5	0.8	0.5	0.0	0.0	0.0	7.1	8.4	5.0	2.9	0.3	0.3	0.0
Home goals	3	5.0	6.8	2.1	0.5	0.5	0.0	0.0	0.0	3.9	3.4	1.1	0.8	0.0	0.0	0.0
	4	2.9	1.3	1.6	0.5	0.0	0.0	0.0	0.0	2.4	2.9	0.3	0.5	0.0	0.0	0.0
	5	0.8	0.5	0.0	0.0	0.3	0.0	0.0	0.0	2.4	0.8	0.3	0.0	0.3	0.0	0.0
	6	0.0	0.5	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.0

Source: *Soccerbase.com*

TABLE 3: Implied frequency (probability) from average bookmaker odds for scoreline outcomes in the 2016–17 and 2017–18 EPL seasons.

		2016–17 Away goals								2017–18 Away goals							
		0	1	2	3	4	5	6	7	0	1	2	3	4	5	6	7
	0	8.8	7.6	4.0	1.7	0.9	0.7	0.6	0.3	8.5	7.3	3.6	1.5	0.9	0.6	0.6	0.4
	1	10.5	13.1	6.7	2.5	1.0	0.7	0.3	0.3	10.2	12.6	6.2	2.3	1.0	0.7	0.3	0.4
	2	6.8	9.1	5.9	2.3	0.9	0.4	0.3	0.2	6.5	8.7	5.6	2.1	0.9	0.4	0.3	0.3
Home goals	3	3.1	4.2	3.0	1.5	0.6	0.3	0.3	0.1	2.9	3.9	2.7	1.3	0.5	0.3	0.3	0.3
	4	1.4	1.7	1.3	0.7	0.4	0.3	0.2	0.1	1.4	1.6	1.2	0.6	0.4	0.3	0.3	
	5	0.9	0.9	0.6	0.3	0.3	0.2	0.2	0.1	0.9	0.9	0.5	0.4	0.3	0.3	0.3	
	6	0.6	0.4	0.3	0.3	0.2	0.2	0.2	0.1	0.7	0.3	0.4	0.3	0.3	0.3	0.3	
	7	0.3	0.3	0.2	0.1	0.1	0.1	0.1	0.1	0.7	0.6	0.4	0.3				

Source: author calculations using *Oddsportal.com* and *Soccerbase.com*

TABLE 4: Weak & strong efficiency tests for forecast scoreline outcomes

	Weak		Strong	
	Model	Bookmakers	Model	Bookmakers
	(1)	(2)	(3)	(4)
Constant ($\widehat{\alpha}$)	0.002	-0.002	0.002	-0.002
	(0.002)	(0.003)	(0.002)	(0.003)
Forecast ($\widehat{\beta}$)	$0.839^{***‡‡}$	$1.156^{***‡‡}$	$0.839^{***‡‡}$	$1.156^{***‡‡}$
	(0.014)	(0.018)	(0.014)	(0.018)
Scoreline freq.			-0.00005	0.001
			(0.015)	(0.015)
Points (H)			0.00000	0.00001
			(0.00003)	(0.00003)
Points diff.			-0.00000	-0.00001
			(0.0001)	(0.0001)
Form (H)			0.00000	-0.00003
			(0.0002)	(0.0002)
Form diff.			0.00000	0.00001
			(0.0001)	(0.0001)
Elo prediction			0.00001	-0.0001
			(0.004)	(0.004)
Observations	61,560	61,560	61,560	61,560
Adjusted R^2	0.052	0.063	0.052	0.063
Resid. std. error	0.107	0.107	0.108	0.107
Ftest of efficiency	0.000^{***}	0.000^{***}	0.000^{***}	0.000^{***}

Notes: *p<0.1; **p<0.05; ***p<0.01, two-tailed tests of difference from zero. †p<0.1; ‡p<0.05; ‡‡p<0.01, two-tailed tests of difference from one for $\widehat{\beta}$.

TABLE 5: Weak & strong efficiency tests for forecast result outcomes (home win, draw, away win)

	Weak		Strong	
	Model	Bookmakers	Model	Bookmakers
	(1)	(2)	(3)	(4)
Constant ($\widehat{\alpha}$)	0.112***	0.043	0.005	0.071
	(0.040)	(0.038)	(0.045)	(0.045)
Home-win forecast ($\widehat{\beta}$)	0.810***‡	0.957***	0.317**‡‡	1.158***
	(0.080)	(0.076)	(0.130)	(0.200)
Elo prediction			0.660***	−0.238
			(0.138)	(0.215)
Adjusted R^2	0.117	0.173	0.142	0.176
F-test of efficiency	0.919	0.995	0.610	0.978
Constant ($\hat{\alpha}$)	0.116**	0.005	0.195***	0.016
	(0.052)	(0.061)	(0.065)	(0.102)
Draw forecast ($\widehat{\beta}$)	0.482**‡‡	0.979***	0.299‡‡	0.945***
	(0.191)	(0.246)	(0.211)	(0.354)
Elo predict (balance)			−0.795**	−0.068
			(0.393)	(0.508))
Adjusted R^2	0.007	0.019	0.011	0.020
F-test of efficiency	0.894	1.000	0.835	1.000
Constant ($\widehat{\alpha}$)	0.023	−0.047*	0.432***	−0.313**
	(0.028)	(0.027)	(0.091)	(0.131)
Away-win forecast ($\widehat{\alpha}$)	0.892***	1.090***	0.442***‡‡	1.406***‡
	(0.079)	(0.074)	(0.124)	(0.169)
Elo prediction			−0.557***	0.343*
			(0.119)	(0.165)
Adjusted R^2	0.143	0.221	0.166	0.225
F-test of efficiency	0.973	0.979	0.67	0.916
Observations	760	759	760	759

Note: *p<0.1; **p<0.05; ***p<0.01, two-tailed tests. †p<0.1; ‡p<0.05; ‡‡p<0.01, two-tailed tests of difference from one for $\widehat{\beta}$.

TABLE 6: Encompassing testing for scoreline forecasts

	Model Prob.	Bookmaker
Model Prob.		1.80
Bookmaker	**8.77**	

Note: bold-faced numbers indicate t-statistics larger than 3.8, i.e., significant values, based on the rule of thumb established in Campos et al. (2003) for adjusting t-statistics with large sample sizes. The positive sign of the statistics implies that the column forecasts on average increase the errors of the row forecasts.

TABLE 7: Applying a simple betting strategy using the scoreline forecasting model

	ROI (%)			
	Result (1)	Scoreline (2)	Margin (3)	Total Goals (4)
2016/17				
Unconditional	-3.4	-10.8	-3.9	-1.9
Conditional	12.7	-5.2	3.1	-1.6
2017/18				
Unconditional	4.8	-25.8	-6.7	-6.5
Conditional	-0.2	-26.6	-6.5	-7.6

Notes: Columns (1)-(4) give implied returns on investment from betting the amount x over the whole season on each and every match, consistent with the scoreline point forecast made based on the statistical model (row), i.e., a total investment by the forecaster/bettor over the season of $380x$ for either the result, scoreline, margin or total number of goals in a match.